**INFERENCES**

**>** prior\_summary**(**fit**)**

prior class coef group resp dpar nlpar lb ub source

normal**(**0, 10**)** b user

normal**(**0, 10**)** b alcohol **(**vectorized**)**

normal**(**0, 10**)** b body **(**vectorized**)**

normal**(**0, 10**)** b price\_log10 **(**vectorized**)**

normal**(**0, 10**)** b price\_log10**:**body **(**vectorized**)**

normal**(**0, 10**)** b price\_log10**:**tfidf\_tsne\_1\_norm **(**vectorized**)**

normal**(**0, 10**)** b Rich **(**vectorized**)**

normal**(**0, 10**)** b tannin **(**vectorized**)**

normal**(**0, 10**)** b tfidf\_tsne\_1\_norm **(**vectorized**)**

student\_t**(**3, 0, 2.5**)** Intercept default

student\_t**(**3, 0, 2.5**)** sd 0 default

student\_t**(**3, 0, 2.5**)** sd variety 0 **(**vectorized**)**

student\_t**(**3, 0, 2.5**)** sd Intercept variety 0 **(**vectorized**)**

INFERENCES

* The model has assumed that the random effect intercept follows a Student's T distribution.

**>** summary**(**fit**)**

Family**:** bernoulli

Links**:** mu **=** logit

Formula**:** superior\_rating **~** 1 **+** price\_log10 **+** tannin **+** alcohol **+** Rich **+** price\_log10 **\*** body **+** price\_log10 **\*** tfidf\_tsne\_1\_norm **+** **(**1 **|** variety**)**

Data**:** wine\_reviews **(**Number of observations**:** 2500**)**

Draws**:** 4 chains, each with iter **=** 2000; warmup **=** 1000; thin **=** 1;

total post**-**warmup draws **=** 4000

Multilevel Hyperparameters**:**

**~**variety **(**Number of levels**:** 35**)**

Estimate Est.Error l**-**95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS

sd**(**Intercept**)** 0.65 0.15 0.41 0.98 1.00 1035 2102

Regression Coefficients**:**

Estimate Est.Error l**-**95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS

Intercept **-**10.55 1.84 **-**14.08 **-**6.88 1.00 2175 2400

price\_log10 4.96 1.19 2.57 7.27 1.00 2153 2356

tannin 2.79 0.72 1.37 4.21 1.00 4438 2946

alcohol 2.72 0.60 1.49 3.92 1.00 4403 3301

Rich 1.02 0.14 0.75 1.30 1.00 4389 2703

body 0.97 2.83 **-**4.75 6.51 1.00 2312 2370

tfidf\_tsne\_1\_norm **-**4.40 1.77 **-**7.95 **-**0.98 1.00 2738 2637

price\_log10**:**body **-**0.56 1.87 **-**4.19 3.23 1.00 2266 2652

price\_log10**:**tfidf\_tsne\_1\_norm 2.67 1.15 0.47 4.96 1.00 2802 2656

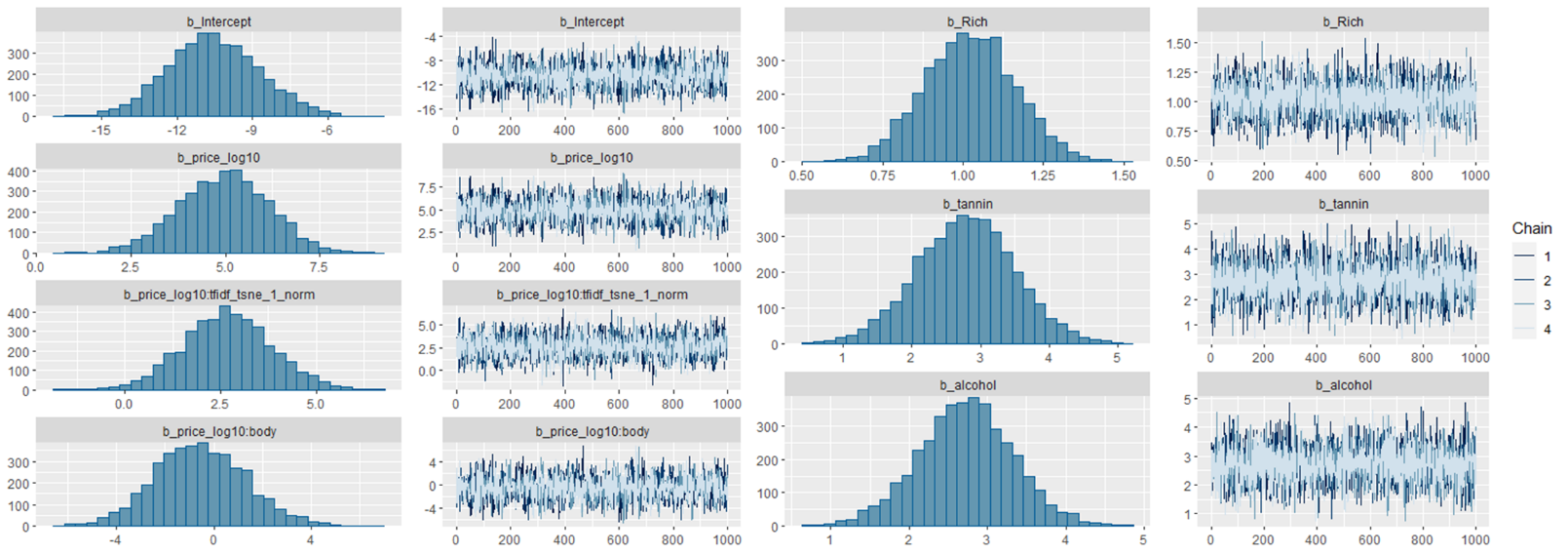
Draws were sampled using sampling**(**NUTS**)**. For each parameter, Bulk\_ESS

and Tail\_ESS are effective sample size measures, and Rhat is the potential

scale reduction factor on split chains **(**at convergence, Rhat **=** 1**)**.

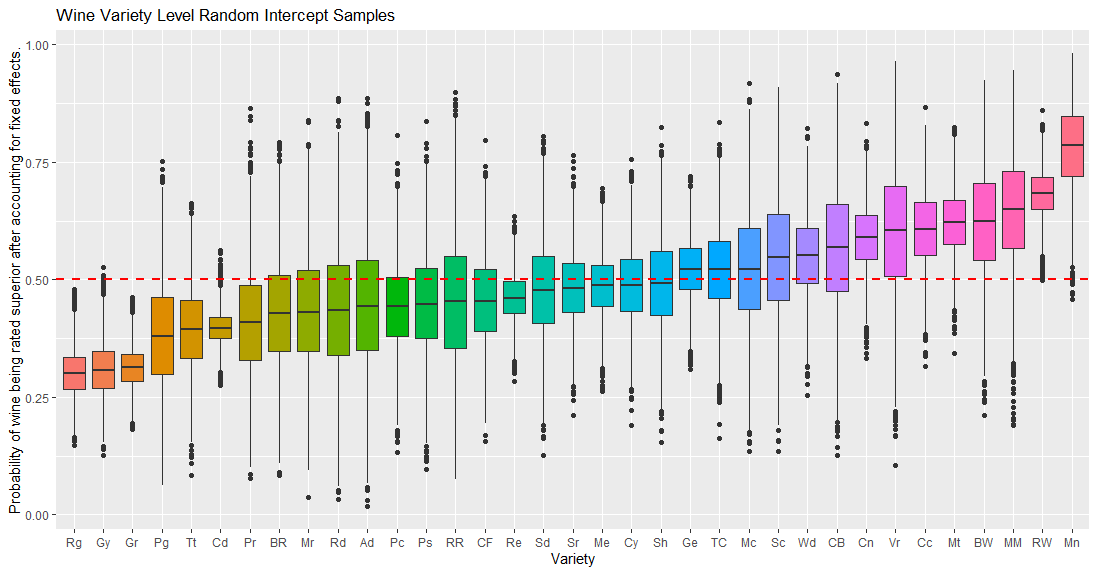
INFERENCES

* *Of all the fixed effects in the model, has the highest average effect* (comparatively highest coefficient of 4.96) on the log-odds of a wine being rated as superior. This was expected given high correlation with response variable from correlation matrix. What this means is that a one-unit increase in price (log transformed) is associated with an increase in the log-odds of wine being "superior" by around 5 units on average (estimate = 4.96) give or take around 1.5 units (.
* The variable with next most noticeable effect is with a negative estimate suggesting a drop in log-odds of success (wine rated superior) with increase in predictor value. Next greatest effect (positive) is asserted by derived features, alcohol, and tannin. The model estimates that a one-unit increase in alcohol/tanning content is associated with an increase of approximately 3 units on average, in the log-odds of wine being rated "superior" (with comparatively least uncertainty ). The fact that these derived features have a larger effect size compared to the single indicator "Rich" reinforces the value of having derived them to represent key wine judgement criteria while incorporating more than one indicator. The interaction term of and however, had least effect (smallest coefficient estimate).
* The model predicts a very low log-odds (Intercept estimate = -10.55) of a wine being rated as "superior" when all other factors are zero. *This hints at a baseline tendency for wines to not get rated as being "superior".* That said, the presence of the random intercept for "variety" with a quite high standard deviation (1.84) gives clues about how some variety-specific effects might be at play.



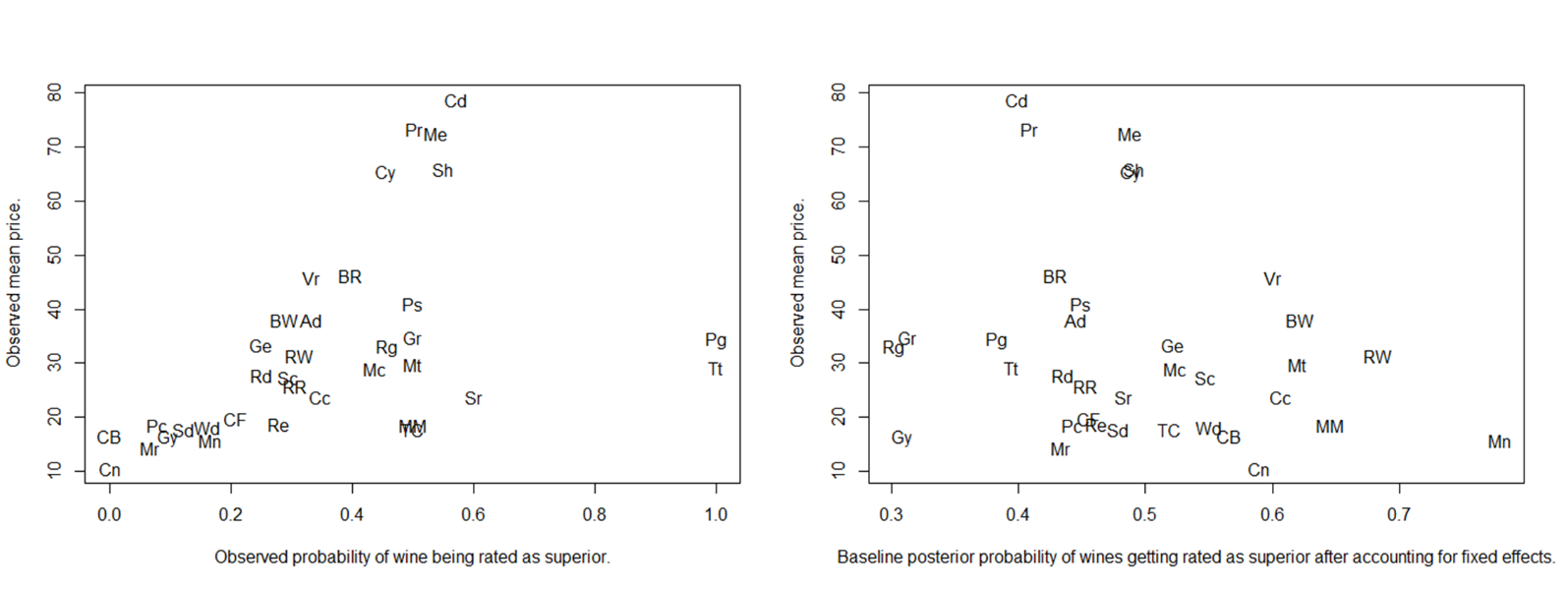
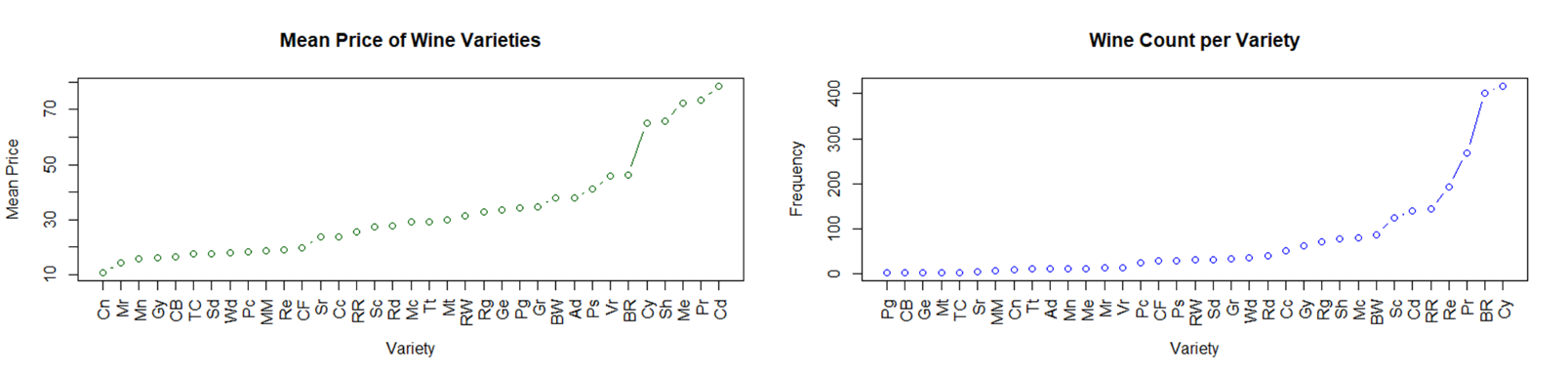
INFERENCES

* Converting from log scale to probability scale, here, a one unit increase in price in the log scale would lead to an increase in the probability of wine being rated "superior" by a factor of anywhere between and times. At first glance, these numbers seem super large. But upon deeper analysis, this is plausible because an increase of 13 – is relative to some baseline and not absolute change. So, if the baseline probability of a wine being rated "superior" was low, say 0.01, even a 1500-fold increase would only be an absolute increase in probability . Because the population parameter intercept was very negative (around -10) in the summary, it might be that baseline probability really is low. Nevertheless, the wide gap (1436 – 13 ) indicates low precision. Perhaps per variety inferences are more precise.
* Generally, if a distribution is centered away from 0 and the credible interval does not have 0, it suggests evidence for a non-zero effect of that variable. That is, the further the distribution is from 0 and the narrower the credible interval, the stronger the evidence. [1] [2] Bearing this in mind, it can be seen here that barring interaction terms, all others seem to be having fair evidence against zero effect. That is, the chance that relationships (positive/negative) observed here between wine characteristics (alcohol, richness, tanning, price) and its likelihood of being ranked as superior, is less likely to be due to random chance.



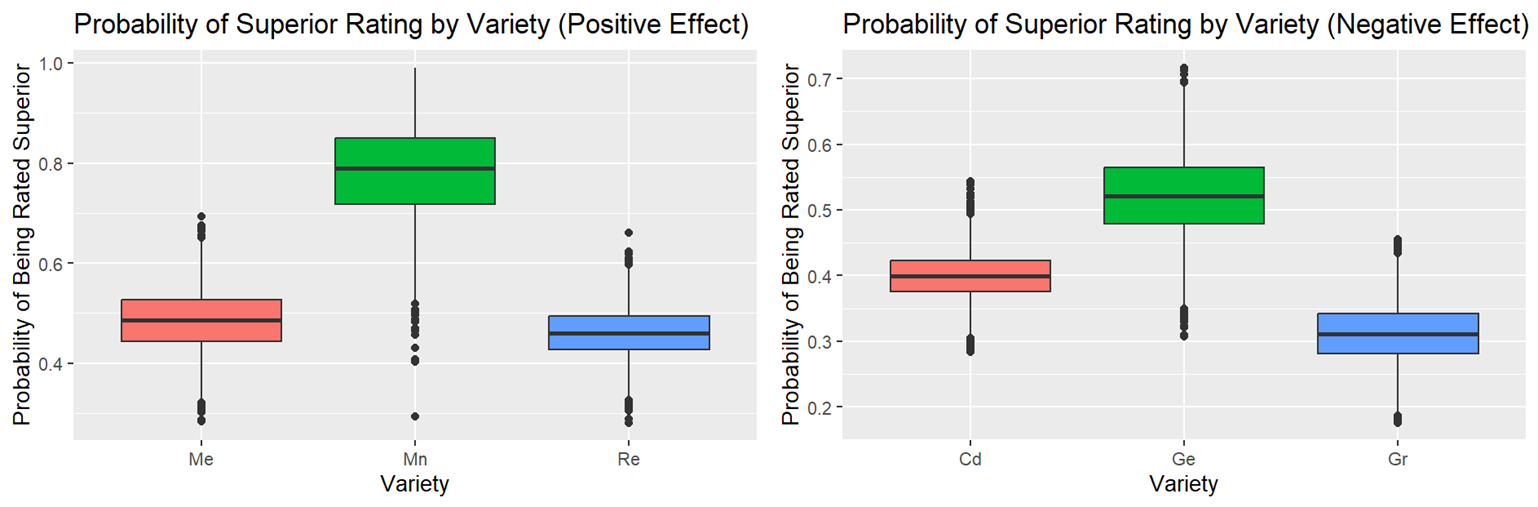
INFERENCES

* Each box in the plot represents the distribution of posterior draws for a specific wine variety. Had the intercepts been directly plotted on the y axis, this would have indicated the per-variety baseline log-odds of a wine being rated "superior" after accounting for the fixed effects in the model due to the predictor variables. Since that might be harder to readily infer from, log-odds were converted to probabilities via . Now, the y-axis indicates baseline probability that a wine of a particular variety gets rated as "superior" after accounting for fixed effects in the model due to predictor variables.
* The 3 wine varieties at the leftmost end of the figure) (Riesling, Gamay, Gewürztraminer) are the ones with the lowest probability of having wines be rated as superior with minimal variance (more certainty).
* Varieties before Cn (Cabernet Sauvignon), all have median intercept value below/very close to 0.5, meaning that consequently suggests that the odds of these varieties of wines being rated high enough to be ranked as superior is low.
* The wine variety with the highest odds of being rated as superior is “Melon” on the extreme right of the figure. It’s median, as well as that of neighboring varieties (“Rhône-style White Blend”, “Malbec-Merlot”, “Bordeaux-style White Blend”, …, “Cabernet Sauvignon”) lie quite above the 50% probability and thus have discernably higher odds of being ranked as superior to the varieties associated with the boxplots at the extreme left end. Thus, one may conclude that if one was to try a vine of the “Melon” variety, it’s very likely that it was highly rated. Overall, some wine varieties do indeed get ranked high more often than others with few, in this dataset, having lowest odds ( of being ranked as superior.
* That said, wine ranking is based on extremely subjective views. This likely explains the relatively wide spread of distributions and ample outliers in most cases.



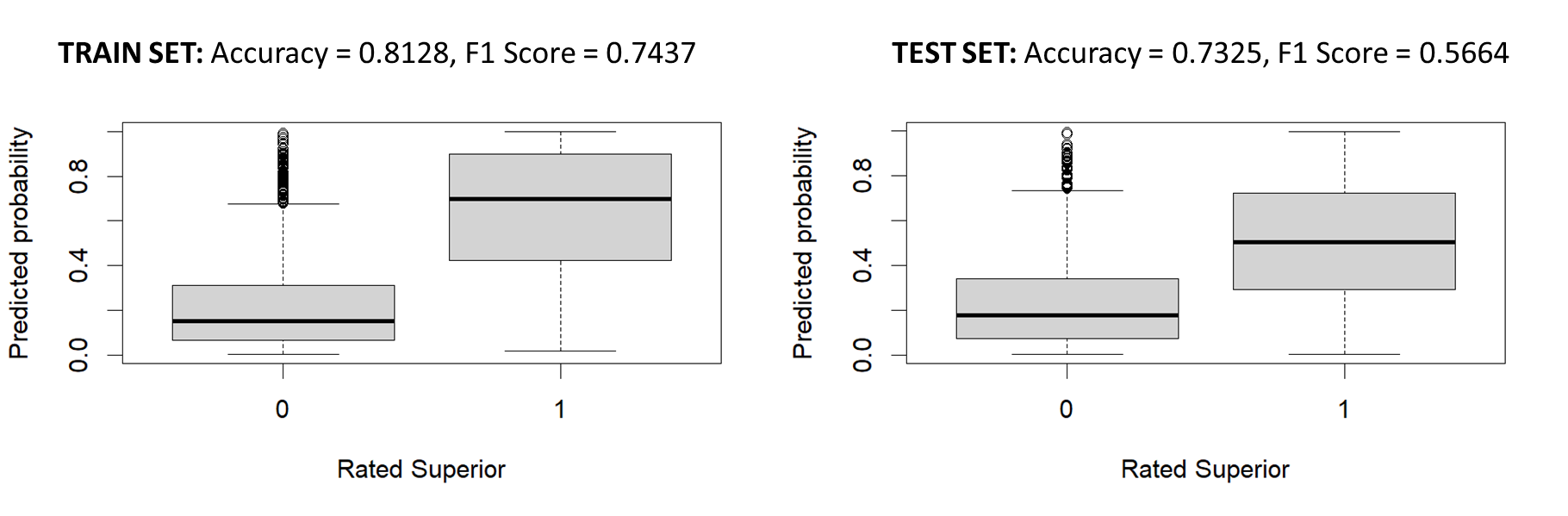
INFERENCES

* Here, subplot 1 shows the unadjusted relationship between price (y axis) and the observed probability of superior rating for each wine variety (x axis). Subplot 2 shows the observed mean price (y axis) v/s baseline posterior probability (log odds exponentiated as on x axis) of wines getting rated as superior per variety after accounting for fixed effects.
* Overall, it can be observed that probability of many wines, especially ones with a high price (in subplot 1 dropped in subplot 2. And many wines that had low probability and price in subplot 1, showed increase from that value in subplot 2.
* An interesting observation was that going from subplot 1 to subplot 2, probability of high-priced wines (e.g. Cd = Champagne Blend) dropped while that of cheaper wines (e.g. Mn = Melon) increased. Varieties like "Cn" (Cabernet Sauvignon) and "CB" (Chenin Blanc-Chardonnay) with 0% observed superior wines (subplot 1) might be assigned higher probabilities (55-60%) in subplot 2 because these varieties may have other characteristics than price, that are generally associated with superior wines. This suggests that price might be partially responsible for observed probabilities and it’s possible that some varieties like Cd = Champagne Blend may be overpriced and that *while in its true that price is positively correlated with price, high price does not necessarily mean wine with great properties. So, you don’t have to break the bank to have great wine. There are plenty of affordable options****.*** Here, *the model recommends wine varieties like Melon (Mn), Malbec-Merlot (MM) and Rhône-style White Blend (RW) as great affordable wines* that may be better overall than some of the high-priced wines based on properties like (alcohol, tannins, body).
* Other extreme changes can be observed in the probability of success associated with some varieties like Pg (Petit Manseng) and Tt (Tannat) which had 100% of observed wines rated superior in subplot 1, but in subplot 2 this drastically dropped to probability of a little less than . This is expected to be because the initial high rating might have been partly due to a small sample size or specific preferences of the raters for those particular wines. By accounting for these factors, the model might have adjusted the probability downwards to a more general level for these varieties.



INFERENCES

* The probability of wine being rated as superior is affected by the variety of wine. This chance is low (a little above 30%) if the wine variety was Gr = Grenache, which increases to a little around 80% if the variety is Melon. We can however, see that there I significant amounts of uncertainty in all cases also comparatively lower for varieties Re (Rosé) and Gr (Grenache).



INFERENCES

* In both cases, the F1 score is lower than accuracy suggesting that there might be an imbalance between precision and recall.
* There is some overfitting given than accuracy and F1 scores on the testing set are lower than on the training set.
* Overall, the model is overfitting the training data. It performs well on data it has already seen but struggles to generalize to unseen data in the test set. This can lead to unreliable predictions on new data. Improvements and alternate approaches should be explored to improve prediction performance.